JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

From Authority Enrichment to AuthorityBox. Applying RDA in a Koha environment

Stefano Bargioni[(a)]

a) Pontificia Università della Santa Croce, http://orcid.org/0000-0001-7679-2874

Contact: Stefano Bargioni, bargioni@pusc.it

Received: 13 September 2019; **Accepted:** 5 November 2019; **First Published:** 15 January 2020

ABSTRACT

If used up to their most important recommendations, RDA guidelines lead to a semantic web oriented catalog. Both authority and bibliographic records require to be curated, especially under the point of view of persistent identifiers, connecting entities to relevant external resources, thus boosting the navigability of the retrieved information and finally of the whole catalog.

To dramatically improve these goals, also established records need to be provided with links, URIs, or persistent identifiers.

An automatic approach was studied and applied in two phases to authority records of a Koha catalog.

A new tool was created and added to the detailed display of a bibliographic record in order to show valuable information for each agent having a responsibility.

KEYWORDS

RDA; Koha; Authority files; Persistent identifiers; VIAF; Linked data; OPAC interface.

CITATION

Bargioni, S. "From Authority Enrichment to AuthorityBox. Applying RDA in a Koha environment." *JLIS.it* 11, 1 (January 2020): 175–189. DOI: 10.4403/jlis.it-12595.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

1. Introduction

In 2014, the Library of the Pontificia Università della Santa Croce started working on authority records in order to link them to records of other institutions or projects, especially to the Virtual International Authority File (VIAF).[1] This decision was a consequence of (1) the migration to the open source integrated library system Koha[2] in 2011 and (2) the introduction of Resource Description and Access (RDA) in the URBE Network.[3] Authority records were prepared in a semiautomated way when imported into Koha, and new records were registered following Anglo American Cataloguing Rules 2nd edition (AACR2) and then RDA. RDA was officially adopted by URBE in March 2017. Right after it, URBE became a member of the European RDA Interest Group (EURIG).[4]

The cataloguers' staff of the Pontificia Università della Santa Croce noticed the importance of comparing records with other sources, especially studying how they were registered by national libraries, and used in their websites. For this reason we decided to store identifiers using tag 024 (MARC 21). Our choice was to add VIAF and ISNI (International Standard Name Identifier)[5] identifiers to our records, following the use of some of the most important cataloging agencies. VIAF identifiers were chosen for the prestige of the VIAF project, while ISNI identifiers for the wide range of contributions and its compliance with the International Organization for Standardization (ISO).

2. Adding persistent identifiers

The manual process of adding identifiers to authority records can be a tedious operation, and can slow down the productivity of the cataloguing process. Steps span from accessing one or more websites, pasting the searched name, comparing results with data available in our library, modifying the search if not successful, and copying the identifier. Often, other important data are manually gathered and reported, such as dates, gender, languages used by the author, alternative names.

A way to face this complexity was studied and applied to Koha, looking for the possibility to search, retrieve and save valuable data programmatically. A Koha extension was written and applied to the cataloguing interface, both in the record view page and in the update page. On loading these pages, the browser accesses the ISNI search programmable interface (API)[6] using the personal or corporate name of the record. In a floating window, results are compared with our record, highlighting matches based on the International Standard Book Number (ISBN). In case of no matches on ISBNs, the cataloguer also can visually detect a match through two lists of titles, from ISNI and from our catalog.

[1] https://viaf.org. VIAF gathers the authority files of national libraries and other projects, and merges them into a cluster record that brings together the different names used worldwide for the same entity.

[2] https://koha-community.org. Koha is a Maori name for gift, and not an acronym. It is a 20 years old open source ILS with a strong spread all over the world.

[3] URBE (Unione Romana Biblioteche Ecclesiastiche, http://www.urbe.it) is a network of 18 academic institutions. The Pontificia Università della Santa Croce belongs to URBE.

[4] http://www.rda-rsc.org/europe.

[5] http://www.isni.org/. For a discussion on ISNI, see MacEwan, Angjeli, Gatenby 2013.

[6] http://isni.oclc.nl/sru/.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

In case tag 024 has no occurrences, a button allows to capture both ISNI and VIAF identifiers and report them into two new occurrences, in compliance with MARC 21 rules:[7]

024 7# $2 viaf $a *viaf identifier*
024 7# $2 isni $a *isni identifier* without spaces

The first indicator, set at value "7", states that the source is specified in subfield "2" with a coded value listed in "Standard Identifier Source Codes".[8]
We chose to store in subfield "a" the basic information, i.e. the identifiers instead of the Uniform Resource Identifiers (URI), leaving other applications to rebuild URIs, or build links and services in the simplest way as possible, as described later.
Thanks to this approach, in about five years, 14,100 authority records were supplied with identifiers, for 78,500 connected bibliographic records on a total of about 174,300 (45%). This percentage will dramatically increase at the end of Phase II, as shown below in Section 9.

3. Working on backlogs

The previous task was performed manually also on some established authority records, especially for famous authors, authors related to our University or to the URBE network, in an effort of adding information that is often just within our reach. At the same time, these records were declared as RDA compliant[9] in case tags 024, 046, 670 where present.
However, the ordinary work of cataloguing doesn't allow for a good pace to enrich the full authority catalog. This is why we studied and realized a specific batch procedure. We based this task on our experience on bibliographic records. In the past years, in fact, we performed two other automatic enrichments, the first one to add Dewey codes (Bargioni [*et al.*] 2013) and the second one, following RDA recommendations once again, to add relationship codes and designators (Bargioni 2016) to authors, in subfields 1xx$e and 7xx$e[10] of bibliographic records.
Both enrichments represented a boost for this kind of information in our records. From thereafter, cataloging activity includes their use, since the staff -who directly participated in the processes- is trained to do it.
We consider this *boost effect* very valuable, especially for authority records. This latter sometimes can represent an extra work for cataloguers, especially if an authority department can't be established. Moreover, it also allows to introduce new services for end users, such as the AuthorityBox, discussed in Section 10.

4. Enrichment of records without VIAFid (phase I)

Batch reconciliation of local data with other data from external sources can be defined as a process of ensuring that two sets of records are in agreement. It can be performed in different ways: from

[7] https://www.loc.gov/marc/authority/ad024.html.
[8] https://www.loc.gov/standards/sourcelist/standard-identifier.html.
[9] This compliance is stated in tag 040 subfield "e" set to "rda".
[10] For 111 and 711, $j is the subfield to use.

ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

online matches to offline comparisons with other datasets. The recordset to reconcile with an external dataset usually require a fuzzy matching to match entries; the reconciliation is successful when a unique ID from the external dataset is retrieved and stored in the local record.

We studied four approaches: (1) using OpenRefine,[11] (2) using a full-online query of VIAF data, (3) a full offline procedure, and (4) a mix of offline and online.

The (1) procedure based on OpenRefine can be summarized in these steps: extract records or their valuable data, like heading and local identifier, fill in a OpenRefine project, use the VIAF reconcile function[12] on the heading column, manually resolve uncertain matches on the new column generated by the function, add tag 024 to MARC records and update them on the catalog.

For each name, OpenRefine reconciliation services usually return a group of names for each name, that can require manual fixings.

The (2) online approach[13] consists on querying the VIAF autosuggest service, keeping similar names, trying to detect a match based on ISBNs. If a match is satisfactory, the local record can be enriched and saved. This method continuously connects to the VIAF service, of course with a slow pace to gentle access it. It can be performed using only one name at the time, i.e. a local name like "Iohannes Paulus PP. II, s., 1920-2005." couldn't have a match in VIAF, while one of its alternate names available in our record could (e.g.: John Paul II, Pope, 1920-2005; Jan Pawel II, papiez, 1920-2005; Giovanni Paolo II, papa, 1920-2005; Wojtyla, Karol, 1920-2005;…). This is due to local variants, applied to ancient and religious authors, following a cataloging principle adopted by the URBE network from 2009, in compliance with AACR2 and in partial accordance with the Vatican Library.

Thus we studied the (3) offline approach, to avoid a high number of connections to the VIAF service. VIAF offers its big data in a variety of formats in the Data Source page.[14] The dataset containing all clusters[15] is a very large file especially when it's expanded, loaded and indexed in a database system: too much for our organization.

This is why we chose and applied the (4) partial offline solution: we stored locally only headings, ISBNs and dates from the dataset of VIAF clusters, to perform matches offline, and retrieve the full cluster information only when a match is detected. This solution allowed also to work on authority records one at a time, and perform the batch process during the daily cataloguing work and the online public access to the system. Moreover, we also avoided extracting, updating, importing and reindexing a large amount of authority records, a procedure that requires a period of inactivity of the system.

[11] http://openrefine.org.

[12] OpenRefine, starting from version 2.8, comes with two preinstalled reconciliation services, VIAF and Wikidata. A tutorial https://github.com/OpenRefine/OpenRefine/wiki/Reconciliation-Service-API allows to develop reconciliation services for other sources. https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine illustrates how to use OpenRefine for Wikidata reconciliation.

[13] Two examples of this method are: https://gist.github.com/nichtich/832052 by J. Voss (2010), and http://infomotions.com/blog/2016/05/viaf-finder/ by E. L. Morgan (2016). Both accessed September 10, 2019.

[14] http://viaf.org/viaf/data.

[15] http://viaf.org/viaf/data/viaf-20190203-clusters-marc21.iso.gz. VIAF regenerates the dataset page regularly, on a monthly basis. This link is now invalid, since datasets are replaced monthly.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

5. The reconciliation process in detail

The VIAF dataset of clusters was downloaded[16] and expanded. From a download size of 10.39 gigabytes, the resulting MARCXML records occupied 33.46 gigabytes.

A Perl script[17] was written to extract VIAF identifiers, names, ISBNs and dates.[18] A SQLite[19] database was defined with these columns and filled in with these data, in the same Koha server. The database consisted of 59,168,458 rows. Its first two columns were indexed, rising the size of the database up to 5.1 gigabytes, and forming something similar to a knowledge base. Despite its size, SQLite always performed very quickly, avoiding in this way to set up a DBMS server dedicated to this time limited task.

At the end of the preparation phase, a Perl script was written to run in Koha. Its task can be resumed in the following steps:

- from the Koha catalog, select authority records to be enriched, i.e. lacking tag 024
- for each selected authority
    - extract rows from the SQLite database, starting from names in tags 100$a and eventually from 400$a (alternative names)
    - retrieve their dates from tag 046 or 100$d
    - if a match on dates occurs, enrich the selected authority
    - otherwise, retrieve associated ISBNs from bibliographic records
        - if a match on ISBNs occurs, enrich the selected authority
    - log the results of the operation for tracing and statistical purposes
- end of the process.

"Enrich the selected authority" is the core task. It consists of accessing the VIAF cluster through the VIAFid detected in the match, retrieve the cluster XML record,[20] extract valuable data and update the local authority record.

For valuable data we selected:

- persistent identifiers, other than the VIAFid, like lccn, wikidata, isni, idref, uri (dnb or bnf), bnfcg, datoses, vatlib, nukat,[21] registered in tag 024;
- dates of birth and death, in tag 046;
- gender, in tag 375;
- languages, in tag 377;
- links to Wikipedia pages, in tag 856.

[16] VIAF offers its clusters in a variety of formats, as well as other datasets, updated about once a month.

[17] We adopted Perl for this project because Koha and its libraries are written in this language. Perl libraries used: DBI, XML::LibXML, Business::ISBN, MARC::Record, JSON, List::MoreUtils, and C4::AuthoritiesMarc to read and write authority records in the Koha environment.

[18] While ISBN based matches can be intuitive, libraries can also use date matches to identify people. For a discussion, see Toves and Hickey 2014.

[19] SQLite https://www.sqlite.org, as described in its Wikipedia page https://en.wikipedia.org/wiki/SQLite, "is a relational database management system (RDBMS) contained in a C library. In contrast to many other database management systems, SQLite is not a client–server database engine. Rather, it is embedded into the end program", or it can be used standalone.

[20] A VIAF cluster can be downloaded using the URL https://viaf.org/viaf/VIAFid/viaf.xml.

[21] These source codes of national libraries and international projects, where chosen as a base for new services – usually links – we could add to the OPAC in the future, and especially to greatly enrich our records if published in a LOD format.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

A private note was also added in tag 667 to log the update.
This reconciliation process is quite similar to others applied elsewhere, like the Share-VDE project[22] that works on a *knowledge base* too. It is also quite similar to the process described for instance by Manzotti 2010 to explain the algorithm used by OCLC to group authority records in VIAF clusters.
Our approach tried to minimize the duration of the process, the impact on our server and on the VIAF server, and maximize the accuracy of each match. This is why we preferred to avoid a third level of match, i.e. a match on titles, whose results are hardly ever exact, thus requiring the use of a threshold of similarity that could always be questionable.

6. Some statistics of the reconciliation process

clusters in the VIAF db	26,064,385	
authority records to reconcile	74,931	
records not found in VIAF db	10,942	14.60%
records not reconciled	30,710	40.98%
records reconciled	33,279	44.41%
records reconciled through a date match	576	1.73%
records reconciled through a ISBN match	32,703	98.27%

Table 1. Reconciliation process statistics

Table 1 shows that a large number of authority records were enriched, most of them through a ISBN match. Records not found represent records that are not available elsewhere but in our catalog, especially due to publication related to master degrees at our University and others URBE institutions.
Regarding the quality of the match process, we consider that it was very successful. Very few records were linked to a wrong cluster, and we usually noticed that the error depended on some kind of difficulty faced also by VIAF in its clusterization process due to very common names and surnames, especially of Spanish or English/USA authors.
About persistent identifiers, each recorded in separated occurrences of tag 024, we observed this distribution (Table 2):

sources (024 $2)	**occurrences**	%	**occ/rec**
uri (dnb or bnf)	40,944	19,37%	1,23
viaf	33,279	15,74%	1,00
isni	30,623	14,48%	0,92

[22] http://www.share-vde.org/sharevde/clusters?l=en.

ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

lccn	29,825	14,11%	0,90
idref	26,001	12,30%	0,78
nukat	19,799	9,36%	0,60
wikidata	11,953	5,65%	0,36
vatlib	10,746	5,08%	0,32
datoses	6,958	3,29%	0,21
bnfcg	1,304	0,62%	0,04
TOT occurrences	**211,432**		

Table 2. Distribution of sources of new persistent identifiers

Of course, the VIAFid was always added, while ISNI represents the second important identifier. Note that for tags different from tag 024, we didn't add the information if it was already present. This means that the following figures and statistics refer to new tags.
Figures about dates in tag 046 are shown in Table 3.

Dates added (tag 046)	26,910
Dates of birth added (046$f)	26,742
Dates of death added (046$g)	7,863

Table 3. Dates of birth (subfield f) and death (subfield g)

Tag 046 has the form:

046 ## $2 iso8601 $f yyyy[mm[dd]] $g yyyy[mm[dd]]

where the month and day can be unspecified if unknown. Different precisions of dates is not the main issue about this group of data: despite the effort of the date-processing procedure used by VIAF[23] to assign birth and death dates to a cluster, they are questionable and problematic especially for uncertain dates of ancient or little-known authors. Each cluster in VIAF representing a person has a date range stored as two dates (min and max) consisting of a year, month and day. Together with the date range there is an indication of what they represent: *lived* (the dates are birth and/or death dates), *flourished* (the dates show when the person was active), *circa* (the dates are approximate). Even if we limited the import to type *lived*, we detected some issues that suggest to apply a special attention, up to reject this information. Here are some examples of wrong *lived* dates (as of September 2, 2019):

[23] Date management in VIAF is widely described in Toves and Hickey 2014.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

name	VIAF	birth date	death date
Aubert, Paul	272049937	19480704	1950
Ross, Alex	73992084	19700122	1950
Moltmann, Jürgen	108285879	1926	1950, but he still lives
Albert von Siegburg	68169738	1456	1456

Table 4. Examples of wrong dates in VIAF clusters (as of September 10, 2019)

Figures about gender are listed in Table 5.

Tag 375, subfield a	24,986	
uomo	20,930	83.77%
donna	4,056	16.23%

Table 5. Gender tag 375 added, by value

Figures about languages are listed in Table 6. They were added to 24,565 records, and – of course – more than one occurrence could be added in a record.

eng	8,634	33.43%
ita	4,809	18.62%
ger	3,565	13.80%
fre	3,502	13.56%
spa	2,874	11.13%
lat	587	2.27%
dut	300	1.16%
heb	179	0.69%
grc	165	0.64%
pol	143	0.55%
other languages	1,066	4.13%
Total number of occurrences	25,824	

Table 6. Languages added to tag 377

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

Tag 377 has the form:

377 ## $2 iso639-2
$a language code
$l language term in Italian

Table 7 lists the number of links to Wikipedia added to tag 856, like in

856 4# $u https://en.wikipedia.org/wiki/Pope_John_Paul_II

A total of 20,618 occurrences were added to 10,101 records.

English	en.wikipedia.org	5,711	27.70%
French	fr.wikipedia.org	3,534	17.14%
Italian	it.wikipedia.org	3,076	14.92%
German	de.wikipedia.org	5,480	26.58%
Spanish	es.wikipedia.org	2,817	13.66%

Table 7. Links to Wikipedia pages

We chose to limit the links to the languages mainly used in our University, since VIAF clusters usually contain links to Wikipedia in many other languages.

7. Enrichment of records with VIAFid (phase II)

Records with the VIAFid in 024 were registered during the past years without systematically adding other information. The missing information, and more, can be retrieved from the corresponding VIAF cluster. The process described in Section 4 was performed through a Perl script based on a software library whose objects and methods were reused for this second phase of the project.
A total of 15,274 records to enrich were selected searching for the presence of tag 024 containing a VIAFid, but of course not involved in the previous enrichment phase. For each of them, the corresponding VIAF cluster was downloaded, in XML format.[24] Tags were added if not present, or even enriched when already present:

024	Persistent identifiers other than VIAFid
040	Description conventions
046	Dates of birth and death, including month and day when available

[24] Again using the URL https://viaf.org/viaf/VIAFid/viaf.xml.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

375	Gender
377	Languages used by the author
856	External links to some Wikipedia pages

Table 8. Tags added to records with VIAFid

Tag 040 was present in all modified records, stating the original cataloging agency, the language of cataloging, the transcribing agency; subfield "e", that declares the Description conventions,[25] was set to the value "rda" only if at least fields 024, 046, 670 were present after the enrichment phase.

8. Statistics about the enrichment of records with VIAFid

This phase II involved, as said before, 15,274 authority records. While 15,155 were modified, only 119 weren't modified due to a lack of interesting information in the VIAF cluster. For 235 of them was also possible to correct the VIAFid since in the meanwhile the VIAF data processing modified their cluster ids.[26]
This enrichment phase was applied not only to personal names, but also to any kind of authority record (except for subject headings):

Personal names	14,737
Corporate names	315
Meeting names	87
Preferred (uniform) titles	14
Geographic names	2
Total	15,155

Table 9. Modified authorities by type

9. Global statistics after enrichments

At the end of phase I and phase II of the enrichment process, figures about authority records can be represented as in Table 10, by type and in accordance with RDA:

[25] It must be a code from https://www.loc.gov/standards/sourcelist/descriptive-conventions.html.
[26] VIAF publics the redirected identifiers once a month, in the Data Source page. The published file http://viaf.org/viaf/data/viaf-YYYYMMDD-persist-rdf.xml.gz shows redirections within the VIAF dataset (in RDF). Even if some VIAF services, like retrieving the cluster in XML format, automatically redirect to the current VIAFid, we think that catalogs have to update periodically their URIs to VIAF clusters, applying the redirects listed in this file.

ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

type	non-RDA	RDA	type total	RDA/non-RDA %
Personal names	84,437	14,934	99,371	15.03%
Corporate names	4,023	255	4,278	5,96%
Preferred titles	3,835	273	4,108	6.65%
Subject headings	2,448	983	3,431	28.65%
Meeting names	1,767	68	1,835	3.71%
Geographic names	27	17	44	38.64%
TOTAL	96,537	16,530	113,067	14.62%

Table 10. Counts and distribution of RDA authority records

At the moment, 48,724 authority records have a VIAFid and they are related to 165,610 bibliographic records as of 175,827. This means that about 94% of our bibliographic records contain at least a link to an authority record linked to the VIAF.

10. AuthorityBox

Discussing about how to improve the use of authority record at an end-user level, we concluded that it was possible to show information from authority records as well as from external sources accessed through persistent identifiers. We wrote AuthorityBox as an extension of our OPAC to add information for each agent involved in a bibliographic record.
We added AuthorityBox in the detail page of bibliographic records. It has the form of an accordion, a group of cards or boxes, each related to an agent. A special box was appended for settings, help and info on this extension. Boxes are loaded asynchronously, and are shown closed except for the first one.



Image 1. An example of AuthorityBox, part of the bibliographic record http://catalogo.pusc.it/bib/95161

Each box (image 1) may contain:

- information from the local authority record:
  - dates, gender, languages, biographical or historical information
- links to local services:
  - digital repository, other bibliographic records, etc.
- links to remote services:

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

    - WorldCat Identities, WorldCat IDNetwork, etc.
- a thumbnail of the author from Wikidata or from a local repository
- links to Wikipedia pages
- the permalink of the authority record.

Two black arrows facilitate the opening or closing of all boxes at the same time. Settings in the last box control the info shown in each component, and the size of icons and picture. Settings are saved in a cookie. A "Help" and an "About" section are also provided.

AuthorityBox seems to be the first of its kind. User functions involved in the use of AuthorityBox are: search catalog, identify an item of interest, view record, see, learn, and easily navigate[27] to more information based authority items. It stretches the definition of a library catalog beyond an inventory list, towards a knowledge tool.

Examples of AuthorityBox:

- http://catalogo.pusc.it/bib/182859 (1 author)
- http://catalogo.pusc.it/bib/95161 (5 authors)
- http://catalogo.pusc.it/bib/88801 (10 authors)

Note that the thumbnail, when available, is retrieved by the browser itself, that queries the SPARQL endpoint of Wikidata[28] to obtain the URL and load the image.

11. Future developments

11.1 Enrichment of new records

The acquired quality level in authority records must be maintained. There are two ways to obtain this goal, in our opinion:

- enrich the new records on a regular basis, say once a day;
- add a tool to Koha in the staff cataloging interface to gather useful information from a VIAF cluster while cataloging a new authority record.

The first solution simply requires to apply to new authority records the same script used in Phase II. This is our current procedure. Anyway, the second solution could ensure a better quality of the record, since the catalogers can immediately verify the new added information.

[27] They resemble the FRBR model. See Barbara Tillet, *What is FRBR? A conceptual model for the Bibliographic Universe* (2005), p. 5.

[28] An example: https://query.wikidata.org/sparql?format=json&query= where the query parameter is the following SPARQL query:

```
SELECT *
WHERE { ?p wdt:P214 "VIAFid" . ?p wdt:P18 ?image }
LIMIT 1
```

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12595

11.2 Update of redirected VIAFids

A maintenance procedure could be applied to authority records, to follow the proceedings of the VIAF clusterization process. The redirect file from VIAF Data Source can be downloaded once a month and new redirects applied to local authority records.[29]

11.3 Preferred Titles

Persistent identifiers added to preferred (formerly, uniform) titles can increase the navigability of the catalog and link our records to records of other main catalogs. This enrichment could introduce more valuable relationships in our Koha catalog, towards the introduction of other LRM[30] based functionalities.

11.4 URIs in Wikidata

A proposal to add an identifier of our authority records in Wikidata was filed in August 2018,[31] and accepted: property P5739[32] is now available. Using the SPARQL endpoint of Wikidata, it is possible to select Wikidata entities that have property P5739 as well the VIAFid in property P214.[33] This means that we can consider to enrich Wikidata entities adding P5739 using the match from VIAFids stored in our records and in Wikidata entities.[34]

11.5 Linked Open Data

The next natural step could be the publication of our bibliographic and authority records in RDF format and contribute to the semantic web at the highest level. This goal will require, first of all, to add URIs to entities in the bibliographic records and publish a static or a dynamic triple store.

12. Conclusions

RDA improves the importance of authority records, thanks to the role it gives to URIs and persistent identifiers associated with entities. Even libraries with low resources can face this task and cooperate to the semantic web with their catalogs. Data from important institutions, statically or dynamically available on the net, allow reconciliation procedures required to process existing records and expose them for powerful services and connections. And "help our users with better access to the rich resources we have to offer them" (Tillett 2016, 22).

---

[29] The file http://viaf.org/viaf/data/viaf-20190505-persist-rdf.xml.gz was about 112 megabytes in length. It contained 7,405,741 identifiers redirected to 4,748,305. Of course, only a little part of them needs to be applied to the local catalog.
[30] IFLA LRM is described in https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf and is implemented by RDA.
[31] See https://www.wikidata.org/wiki/Wikidata:Property_proposal/PUSC_ID.
[32] See https://www.wikidata.org/wiki/Property:P5739.
[33] See https://www.wikidata.org/wiki/Property:P214.
[34] This work was realized for the project SHARE Catalogue and described in Possemato and Forziati 2019.

JLIS.it 11, 1 (January 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12595

References

Bargioni, Stefano, Michele Caputo, Alberto Gambardella, and Luigi Gentile. 2013."Obtaining the Dewey Decimal Classification Number from other databases: a catalog clean-up project", *JLis* vol. 4, n. 2. https://www.jlis.it/article/view/8766, DOI: 10.4403/jlis.it-8766. Accessed June 26, 2019.

Bargioni, Stefano. 2016. "Catalog Enrichment for RDA – Adding relationship designators (in Koha)" https://tinyurl.com/yxtmzxhj (slides), https://tinyurl.com/y3zo27h2 (text). Accessed June 26, 2019.

MacEwan, Andrew, Anila Angjeli, and Janifer Gatenby. 2013. "The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control." *Cataloging & Classification Quarterly* vol. 51, Iss. 1-3. DOI: 10.1080/01639374.2012.730601. Accessed September 10, 2019.

Manzotti, Giulia. 2010. "Analysis and Reflections on VIAF, Virtual International Authority File", *JLis.it* vol. 1, n. 2. https://www.jlis.it/article/view/4520. DOI: 10.4403/jlis.it-4520. Abstract in English, text in Italian. Accessed September 10, 2019.

Possemato, Tiziana and Forziati, Claudio, 2019. "Riuso, interoperabilità, influenza: la cooperazione virtuosa tra i progetti SHARE e Wikidata", http://eprints.rclis.org/34350/. In *La biblioteca che cresce. Contenuti e servizi tra frammentazione e integrazione*, Milano, Italy, 14-15 March 2019. [Conference paper]. Accessed September 10, 2019.

Tillett, Barbara B. 2016. "RDA, or, The Long Journey of the Catalog to the Digital Age", *JLis* vol. 7, n. 2, DOI:10.4403/jlis.it-11643. Accessed September 10, 2019.

Toves, Jenny A., and Thomas B. Hickey. 2014. "Parsing and Matching Dates in VIAF", *Code4lib*. https://journal.code4lib.org/articles/9607. Accessed September 2, 2019.